Neuron-based Multifractal Analysis of Neuron Interaction Dynamics in Large Models

Xiongye Xiao[1], Heng Ping[1], Chenyu Zhou[1], Defu Cao[1], Yaxing Li[1], Yizhuo Zhou[2], Shixuan Li[1], Nikos Kanakaris[1], Paul Bogdan[1]

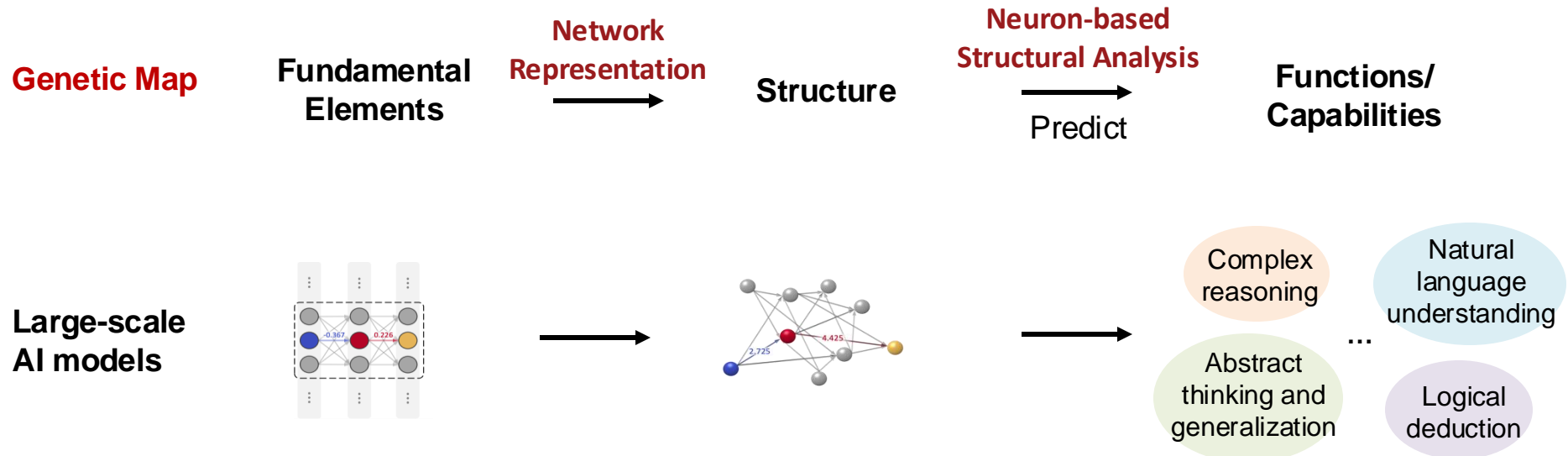[1]University of Southern California, Los Angeles, CA, USA
[2]University of California, Riverside, CA, USA

USC

ICLR

Motivation

## ❑ Sequence-Structure-Function (SSF) principle

SSF principle

Sequence → Predict → **Structure** → Predict → **Function**

How can we develop the "SSF" of non-biological systems?

Genetic Map

Fundamental Elements → Network Representation → **Structure** → Neuron-based Structural Analysis / Predict → **Functions/ Capabilities**

Large-scale AI models

Complex reasoning

Natural language understanding

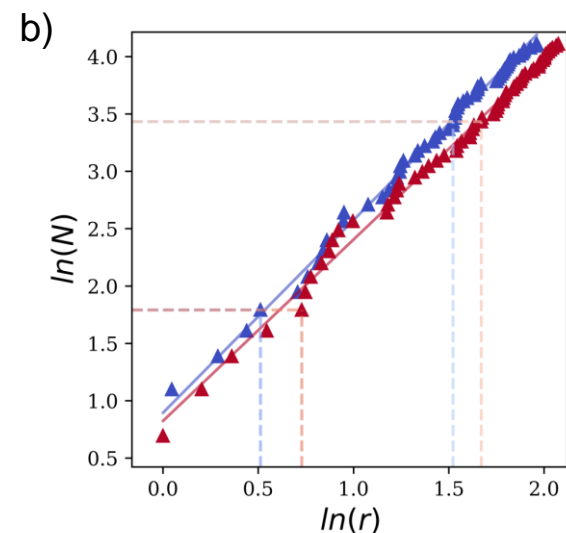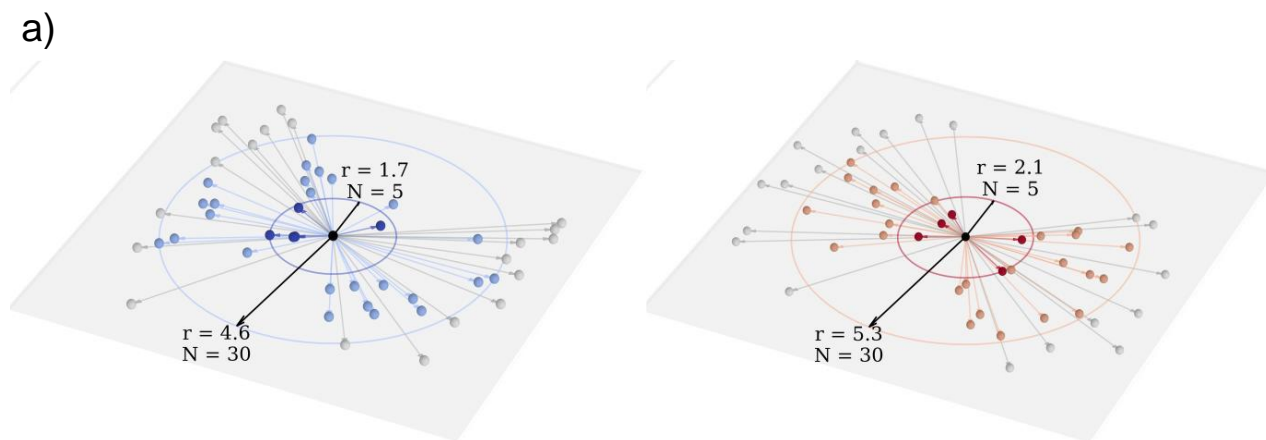Abstract thinking and generalization

...

Logical deduction

(Artificial) Neuronal Interaction Network (NIN)

- **Map artificial neurons and their connections onto NIN**

$$f(w_{ij}) = \begin{cases} \frac{1}{|w_{ij}|}, & \text{if } w_{ij} > \epsilon \\ 0, & \text{otherwise} \end{cases}, \forall i, j \in V$$
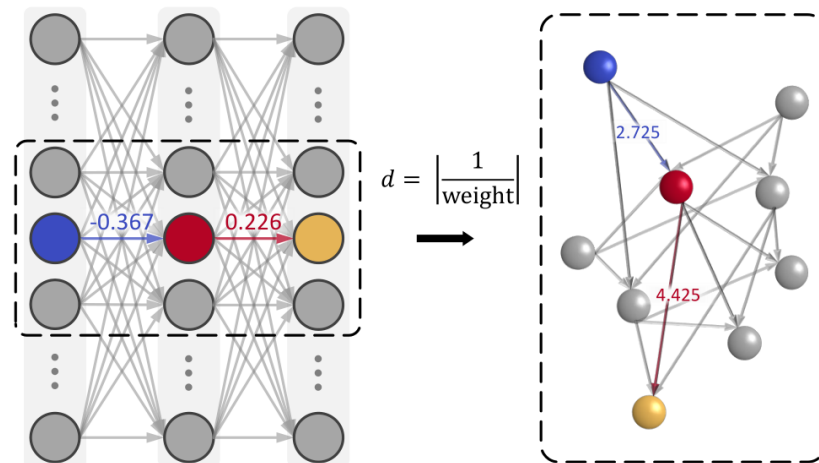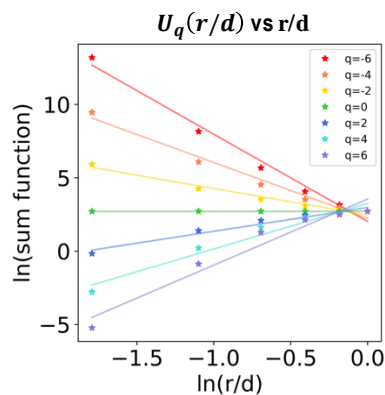

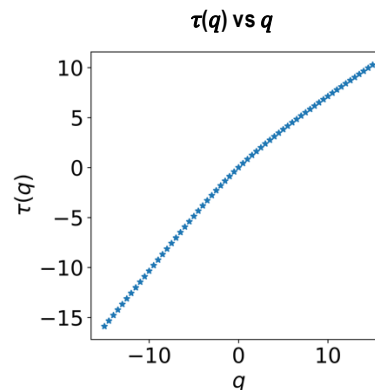- **Fractalities observed in NIN**

3

Neuron-based Multifractal Analysis

❑ **Multifractal Spectrum**

$$\text{Mass Exponent: } \tau(q) = \lim_{l/L \to 0} \frac{\log \sum_{i=1}^{n} \left(\frac{M_i(r)}{M}\right)^q}{\log\left(\frac{r}{d}\right)}$$

$$\text{Lipschitz–Holder Exponent: } \alpha(q) = \frac{d\tau(q)}{dq} = \lim_{l/L \to 0} \frac{\sum_{i=1}^{n}\left[\left(\frac{M_i(r)}{M}\right)^q \log\left(\frac{M_i(r)}{M}\right)\right]}{\sum_{i=1}^{n}\left(\frac{M_i(r)}{M}\right)^q \log\left(\frac{r}{d}\right)}$$
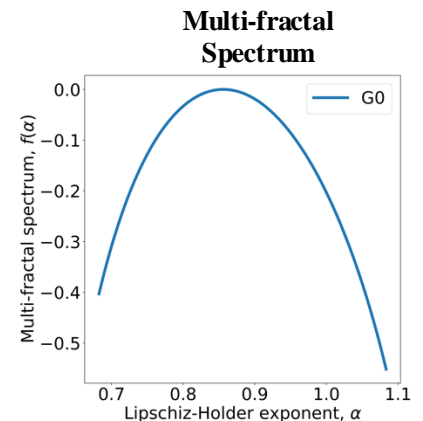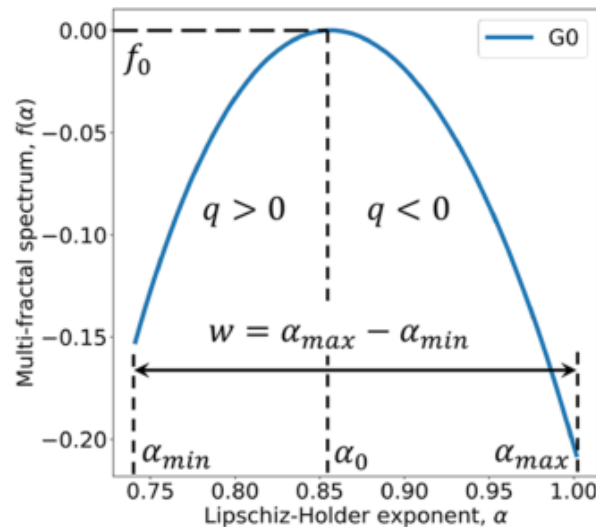

$U_q(r) \sim r^{\tau(q)}$

Legendre transform

$f(\alpha) = q\alpha(q) - \tau(q)$
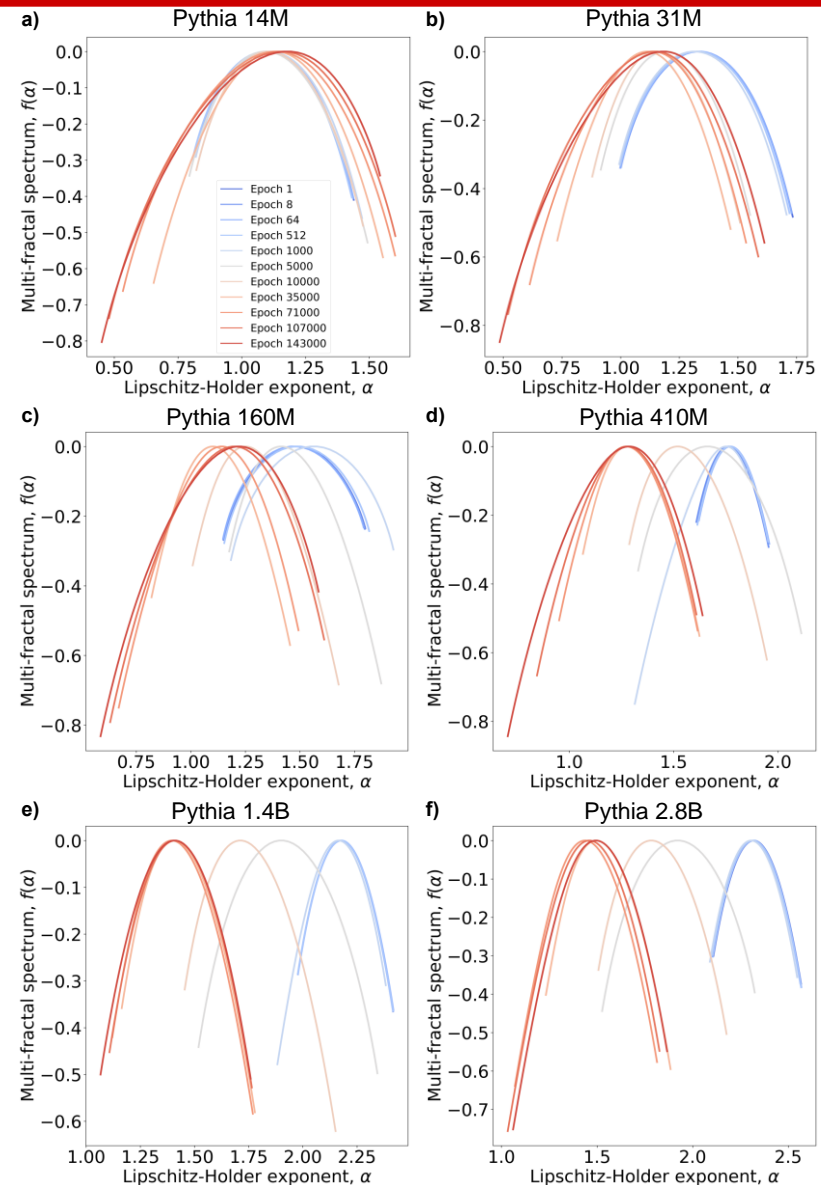
Multifractal Analysis of LLM Training

❑ **Multifractal spectrum**

- ❑ α_0: dominant fractal dimension measuring the degree of irregularity
- ❑ w: degree of heterogeneity

❑ **Self-organization**

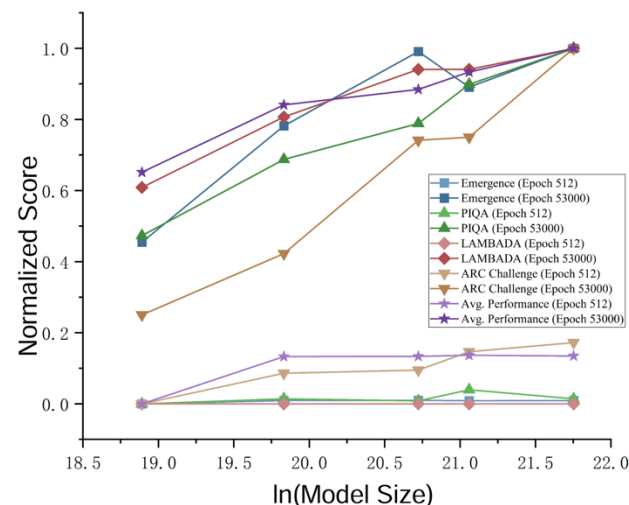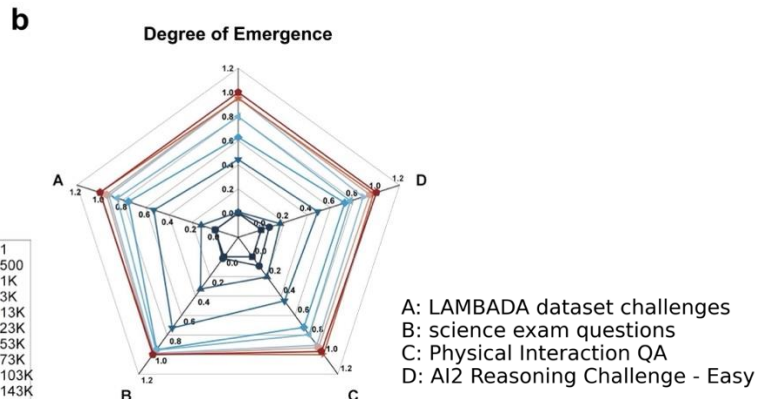- ❑ Increasing heterogeneity (w)
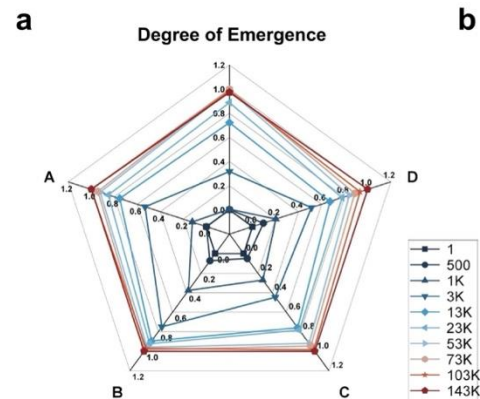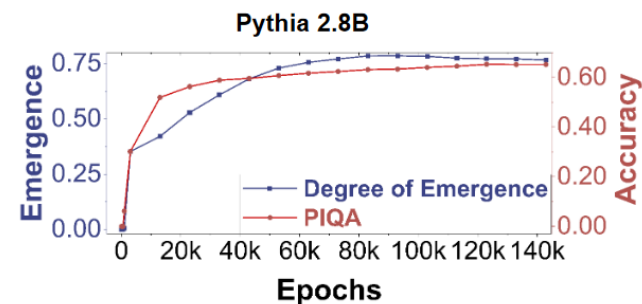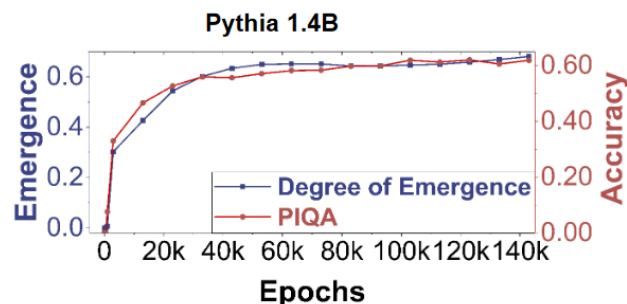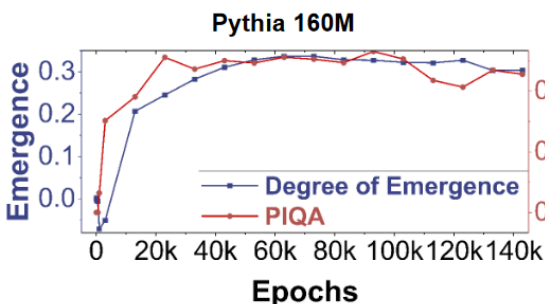- ❑ Increasing regularity (decreasing α_0)

Emergence during Training

☐ **Degree of emergence quantification**

☐ Structural emergence vs. performance

$$E = \frac{w(t)}{w(0)} \log\left(\frac{\alpha_0(0)}{\alpha_0(t)}\right)$$

Relevance Predictability + **New Information**
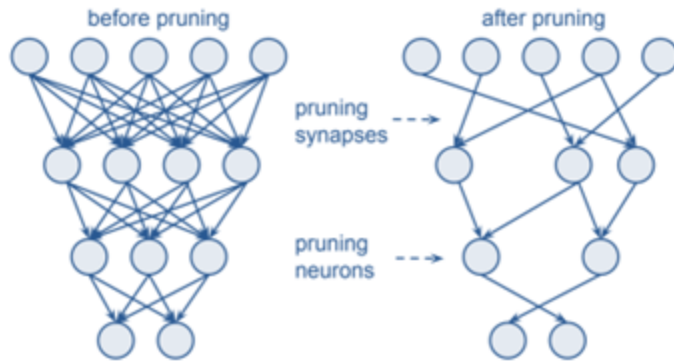

A: LAMBADA dataset challenges
B: science exam questions
C: Physical Interaction QA
D: AI2 Reasoning Challenge - Easy

Science for AI

- ❑ **Applications**

  - ❑ LLM structure optimization

  - ❑ Roguenness/Hallucination detection and correction